The GAMLSS Framework

Modelling skewness and kurtosis

Mikis Stasinopoulos[1]

[1]STORM/LSC, London Metropolitan University

Essex University - May 2013 - Colchester

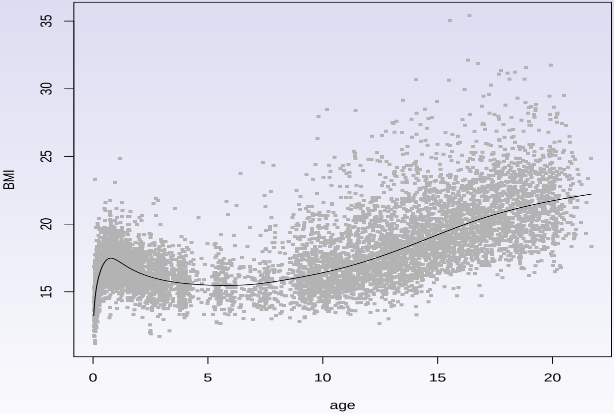
The Dutch boys data

`BMI` : the BMI of 7294 boys

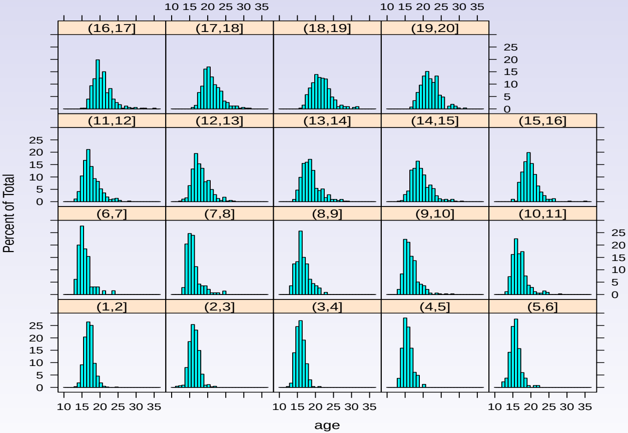
`age` : the age in years

Source: van Buuren and Fredriks (2001)

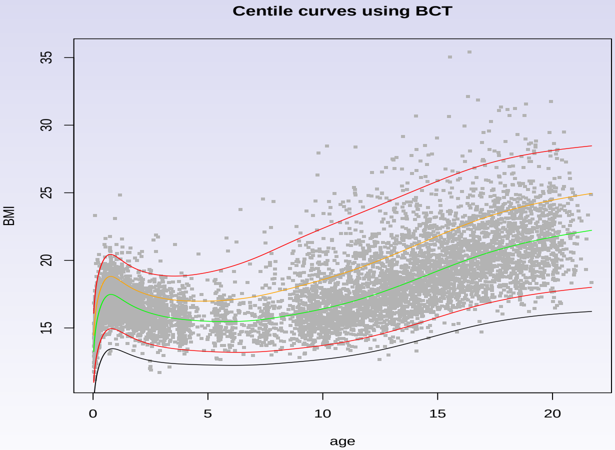
The Dutch boys data: statistical challenges



The Dutch boys data: Histograms by age

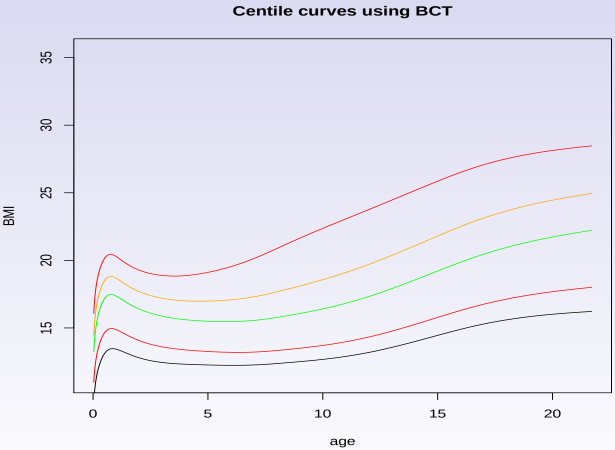


The Dutch boys data: centile estimation



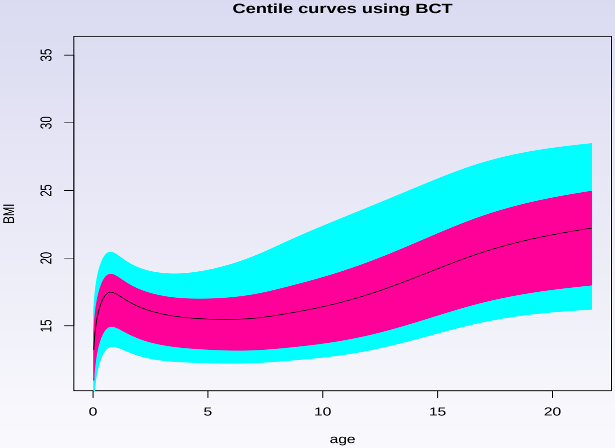
gamlss

The Dutch boys data: centiles

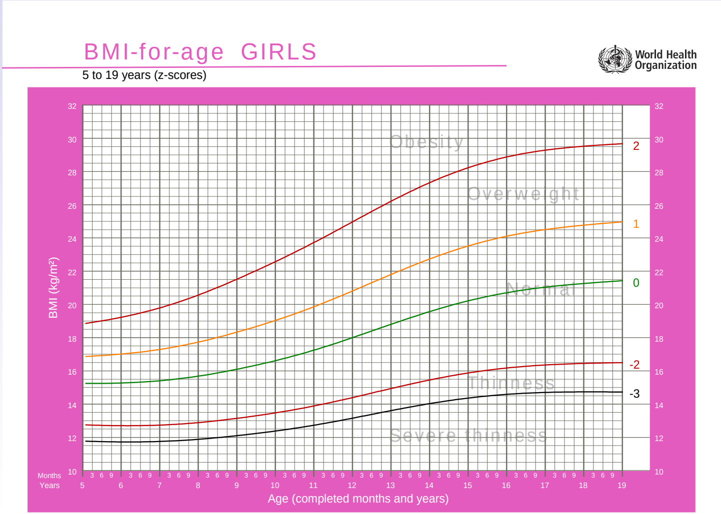


gamlss

The Dutch boys data: fan plot

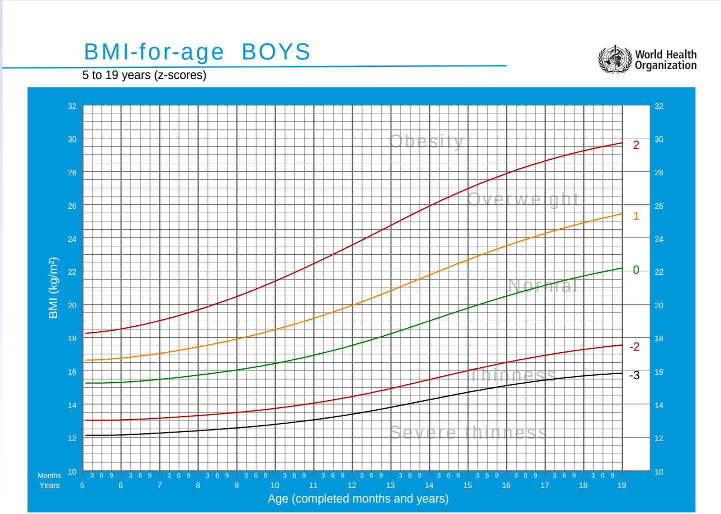


World Health Organisation Child Growth Standards: Girls



gamlss

World Health Organisation Child Growth Standards: Boys



gamlss

The film data

4031 film openings in US 1988-1999:

- `lborev1` : the response variable
  log of the (total revenues - the first week revenue), calculated in 1987 prices
- `lboopen` : the log the first week revenue, calculated in 1987 prices
- `lnosc` : the log of the number of screens
- `dist` : the type of distributor i.e. whether independent or major distributor

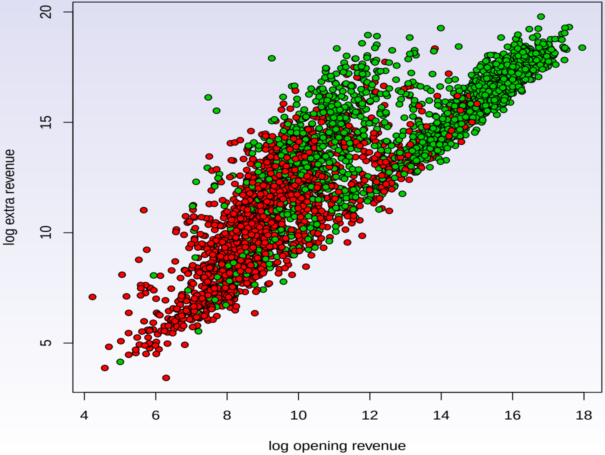
Source: Voudouris *et al.* (2010) Go to Quiz

gamlss

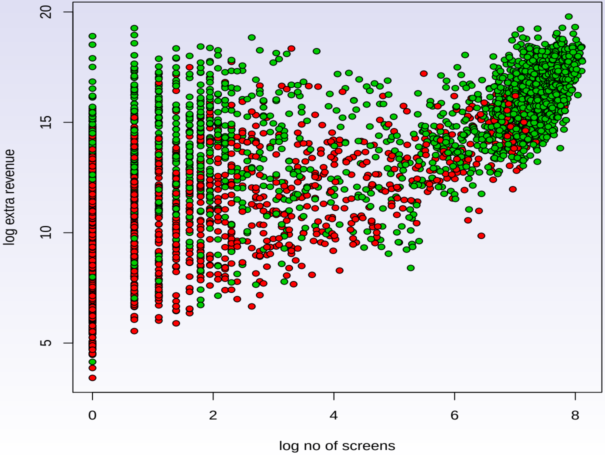
The film data: Question of interest

- Checking whether the "nobody knows anything" principle is valid
- Can we predict the extra revenue given the revenue at the first week?
- Which other explanatory variables are important?
- What is the conditional distribution of the response variable given the explanatory variables?

The film data: log extra revenue against log first week revenue

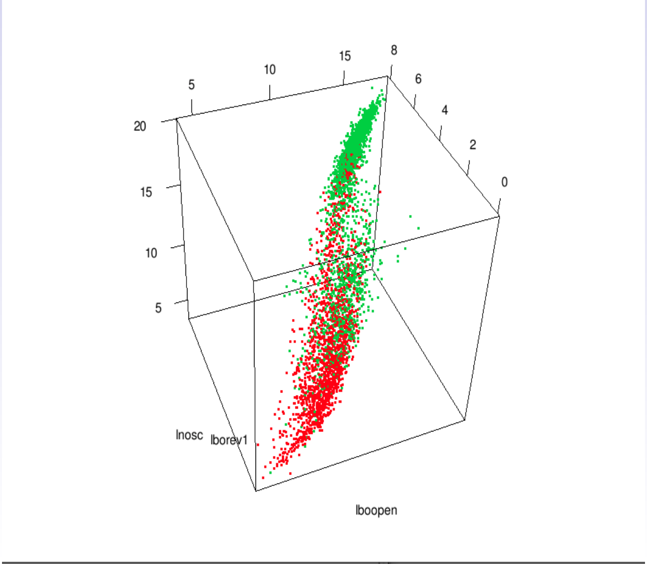


The film data: log extra revenue against log number of screens

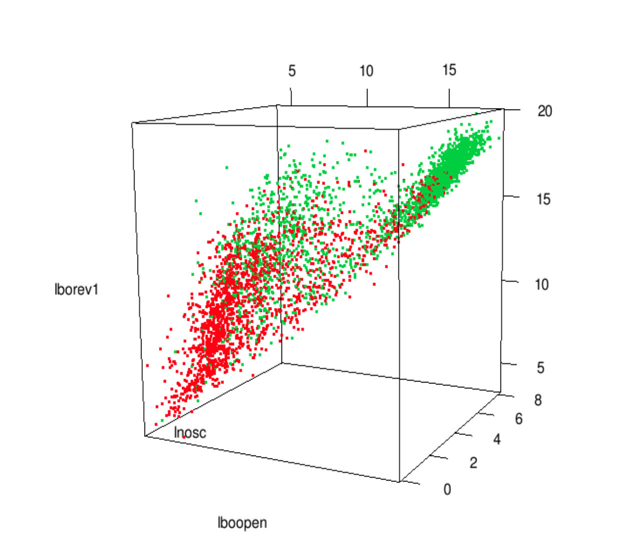


gamlss

The film data: 3d plots

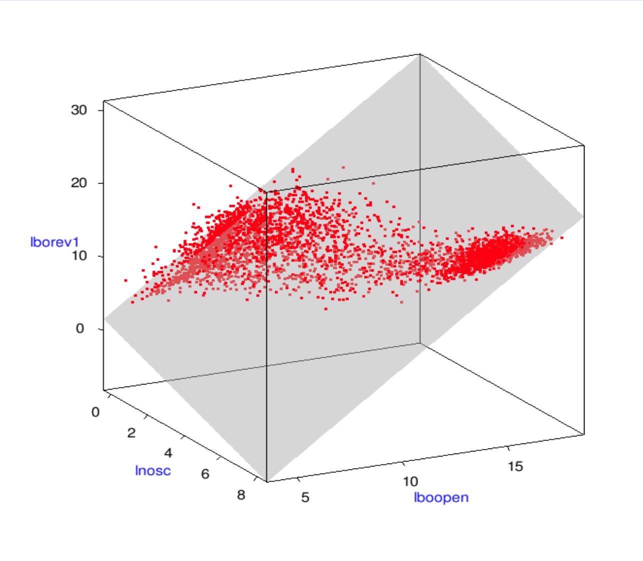


The film data: 3d plots

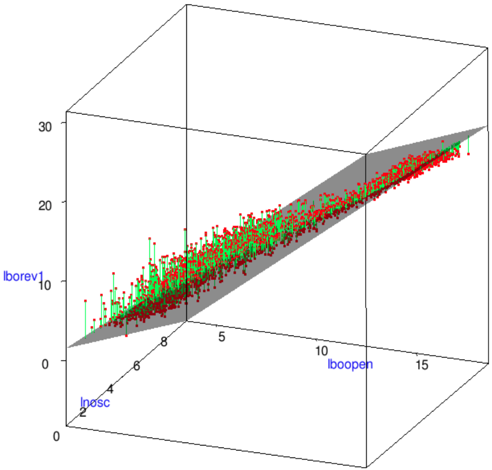


gamlss

The film data: Least Square fit

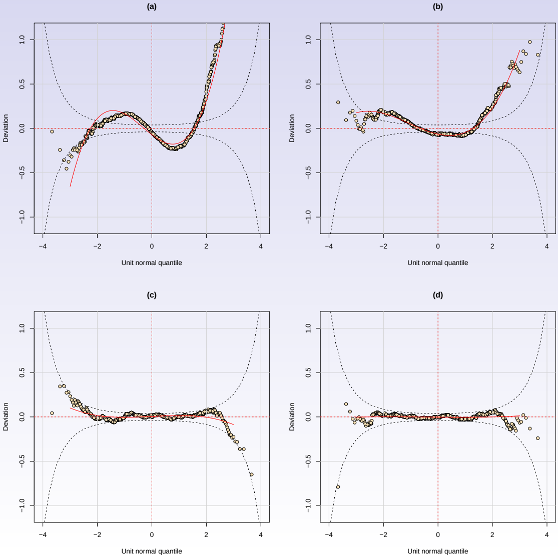


The film data: the residuals from least squares fit



gamlss

The film data: worm plot of the residuals



The Munich rent data

`rent` : the monthly net rent for flats in the city of Munich.

`area` : the floor space area in square meters

`yearc` : year of construction

`location` : whether the location is below, 1, average, 2, or above average 3

`bath` : a factor with levels indicating whether the bath equipment is above average

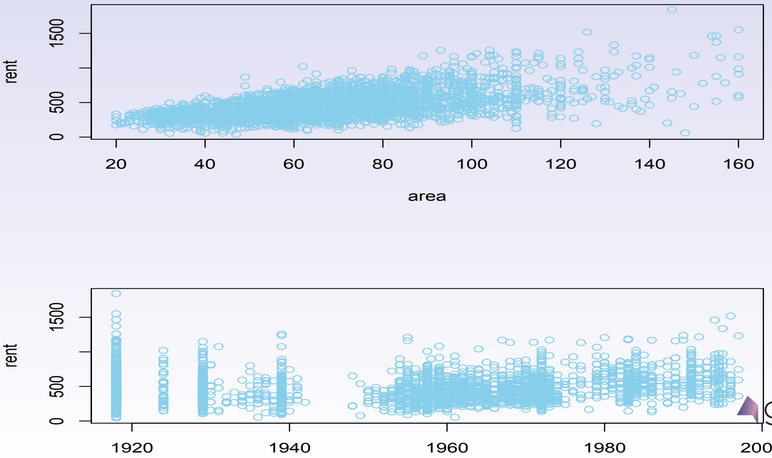
`kitchen` : a factor with levels indicating whether the kitchen equipment is above average

`district` : the district of the flat

Source: Munich rental guide 1999

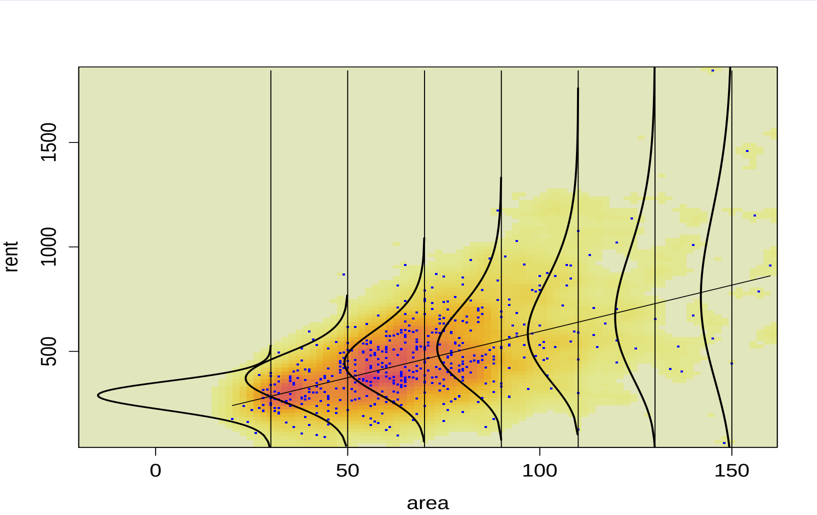
gamlss

The Munich rent data: area and year of construction

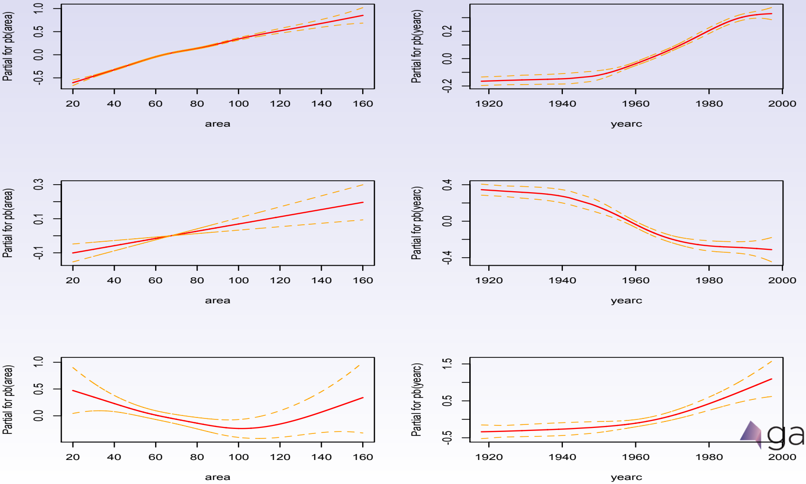


gamlss

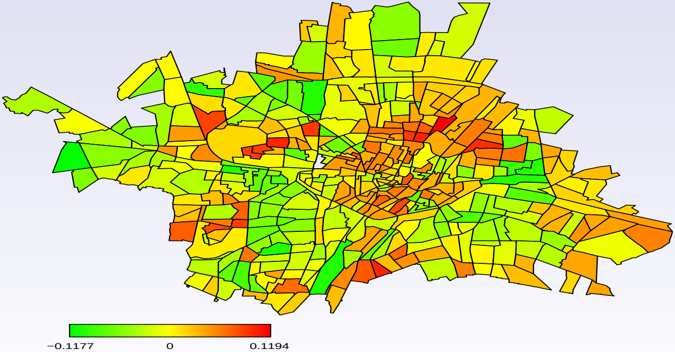
The Munich rent data: area



The Munich rent data: BCCG fitted distribution

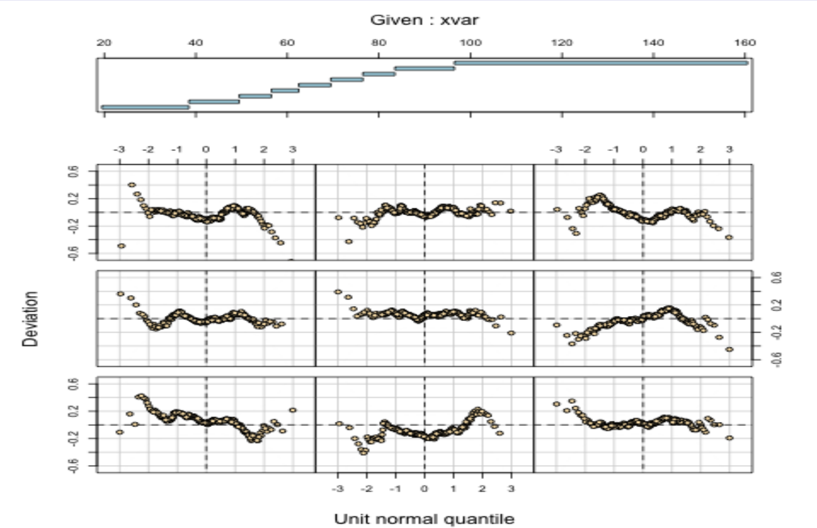


The Munich rent data: district



gamlss

The Munich rent data: worm plots



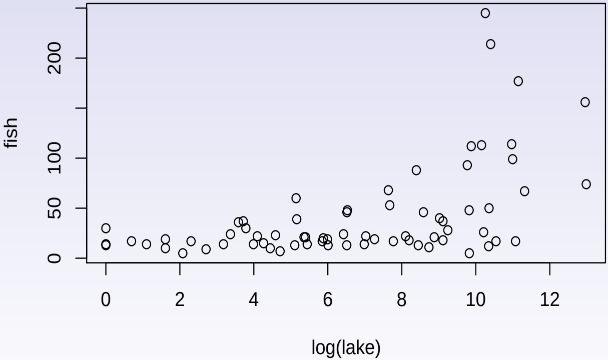
The fish species data

`fish` : the number of different species in 70 lakes in the world

`lake` : the lake area

Source: Stein and Juritz (1988)

The fish species data



gamlss

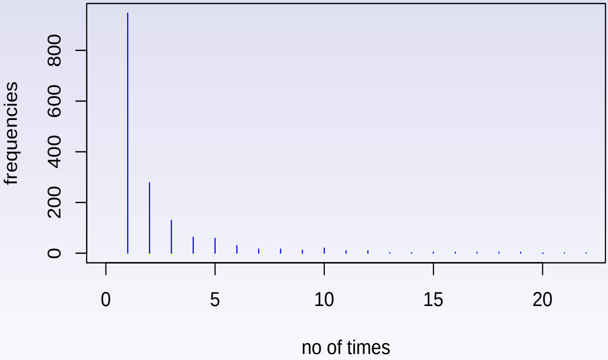
A stylometric application

64 observations

- `word` : is the number of times a word appears in a single text
- `freq` : the number of different words which occur exactly `word` times in the text

Source: Dr Mario Cortina-Borja

The stylometric data



What we need for modelling the above data?

We need

- flexible distributions for the response variable
- to be able to deal with heterogeneity in the data
- to be able to model skewness and kurtosis
- to be able to model overdispersion in count data
- We need modelling all the parameters of the distributions
- flexible functions to model the relationship between the parameter of the distribution and the explanatory variables

The statistical modelling philosophy

Statistical modelling is the art of building parsimonious statistical models for a better understanding of the phenomena of interest.

- Any model is a simplification of reality therefore no model is correct but some of them are useful

- Occam's Razor which states *'entities should not be multiplied beyond necessity'* or KISS (Keep It Simple Stupid)

The statistical modelling philosophy

- Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise. – John W. Tukey (1962), "The future of data analysis", Annals of Mathematical Statistics 33, 1-67 (page 13)

- *"no matter how beautiful your theory, no matter how clever you are or what your name is, if it disagrees with experiment, it's wrong"* (Richard Feynman)

- As statisticians we should try different models and choose the one appropriate for the data

gamlss

What is GAMLSS?

GAMLSS: are semi-parametric regression type models.

- regression type: we have many explanatory variables $\mathbf{X}$ and one response variable $\mathbf{y}$ and we believe that $\mathbf{X} \rightarrow \mathbf{y}$
- parametric: a parametric distribution assumption for the response variable,
- semi: the parameters of the distribution, as functions of explanatory variables, may involve non-parametric smoothing functions
- GAMLSS philosophy: try different models

GAMLSS is a generalisation of GLM and GAM models.

Chronology

Important events in the creation of the GAMLSS models

- up to 1972 Linear model (Gauss) Go to LM
- 1972 Generalised Linear Models (Nelder and Wedderburn) Go to GLM
- 1975 GLIM is released
- 1989 I started working at North London Polytechnic (now London Metropolitan University) where Bob Rigby was working

Chronology 90's

1990 Generalised Additive Models (Hastie and Tibshirani)
Go to GAM

1992 The LMS method of Cole and Green is published
Go to LMS

1993 Mean and Dispersion Additive Model's (MADAM)
Go to MADAM

1999 Creation of Generalised Additive Models for Location Scale and Shape (GAMLSS) Go to GAMLSS

Chronology 2000

2001 First (conference) paper on GAMLSS
The software was transferred from GLIM to R.

2002 GAMLSS paper is send to RSS
WHO contact us about GAMLSS Go to WHO

2004 The GAMLSS paper is read to the Royal Statistical Society Go to Bob Go to Mikis

2004 The GAMLSS software is released to R (CRAN)

2005 The GAMLSS paper is published Go to GAMLSS RSS paper

2006-2012 The GAMLSS impact Go to GAMLSS impact

GAMLSS: Distributions

There are around 80 discrete discrete, continuous continuous, and mixed distributions, mixed, implemented as `gamlss.family` in the R including highly skew and kurtotic distributions Shapes,

- creating a new distribution is relatively easy
- truncating truncated an existing distribution
- using a censored version of an existing distribution
- mixing mixture different distributions to create a new finite mixture distribution.
- discretise discretise continuous distributions
- log or logic any continuous distribution in $(-\infty, \infty)$

Go back 90 next page Additive

gamlss

What is GAMLSS?: additive terms

Additive terms	R Name
Cubic splines	`cs()`, `scs()`
Varying coefficient	`pvc()`
Penalized splines	`ps()`, `pb()`, `cy()`, `tp()`
`loess`/ neural networks	`lo()`, `nn()`
Fractional/power/picewise polynomials	`fp()`, `pp()`, `fk()`
non-linear fit	`nl()`
Random effects	`random()`
Ridge regression	`ri()`
Simon Wood's GAM	`ga()`
random walk and AR	`rw()`, `ar()`

Table: Additive terms implemented within the **gamlss** packages

GAMLSS: R implementation

GAMLSS is implemented in series of packages in R

gamlss	the original package
gamlss.dist	for distributions
gamlss.data	for distributions
gamlss.demo	for demos
gamlss.nl	for non-linear terms
gamlss.tr	for truncated distributions
gamlss.cens	for censored (left, right or interval) response variables
gamlss.mx	for finite mixtures and random effects

The GAMLSS framework packages can be downloaded and installed from CRAN, the R library at http://www.r-project.org/ or at http://www.gamlss.org/.

GAMLSS: model selection

- There are two distinct tasks in selecting a GAMLSS model
  - finding the right distribution
  - selecting the appropriate terms for each parameter of the distribution

GAMLSS: finding the right distribution

- Finding the right distribution is not trivial because it needs good knowledge of the properties of the available distributions.
- We have spend time investigating:
  - the behaviour of the tail of distributions go to tails
  - the skewness and kurtosis properties of four parameter distributions go to properties

GAMLSS: finding appropriate terms

- There are two approaches investigated here:
    - the classical approach based on Generalised Akaike information criterion (GAIC) go to selection
    - boosting go to boosting

gamlss

GAMLSS: present and future

- writing two books
  - on `gamlss.family` distributions
  - on how to use `gamlss` in R
- Is there more time? go to analysis for fish species data
- Agent based models and GAMLSS for modelling energy go to ACEGES
- GAMLSS in C and JAVA
- time series model and GAMLSS go to GEST

Conclusions

- GAMLSS is a very flexible statistical model
- It is a unified framework for univariate regression type of models
- Allows any distribution for the response variable Y
- Models all the parameters of the distribution of Y
- Allows a variety of penalised additive terms in the models for the distribution parameters
- The fitted algorithm is modular, where different components can be added easily
- it can easily introduced to students since it relies on known concepts
- It deals with overdispersion, skewness and kurtosis

This is a collaborative work: A list of people who help with the R implementation

- Bob Rigby (no Bob no GAMLSS)
- Vlasios Voudouris
- Paul Eilers
- Popi Akantziliotou
- Gillian Heller
- Majid Djennad
- Fiona McElduff
- Nicoleta Mortan
- Raydonal Ospina
- Konstantinos Pateras

There are lot more to be done

the END

for more information see

www.gamlss.org

gamlss

The linear model

Linear Model, Gauss

$$\mathbf{y} = \mathbf{X}\beta + \boldsymbol{\epsilon} \text{ where } \boldsymbol{\epsilon} \sim NO(\mathbf{0}, \sigma^2\mathbf{I})$$

The model can be also written as:

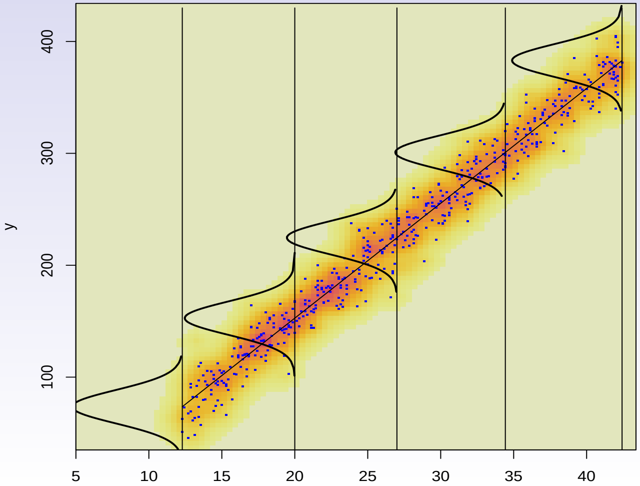
$$\mathbf{y} \sim NO(\boldsymbol{\mu}, \sigma^2\mathbf{I}) \text{ where } \boldsymbol{\mu} = \mathbf{X}\beta$$

Go back 70-80

gamlss

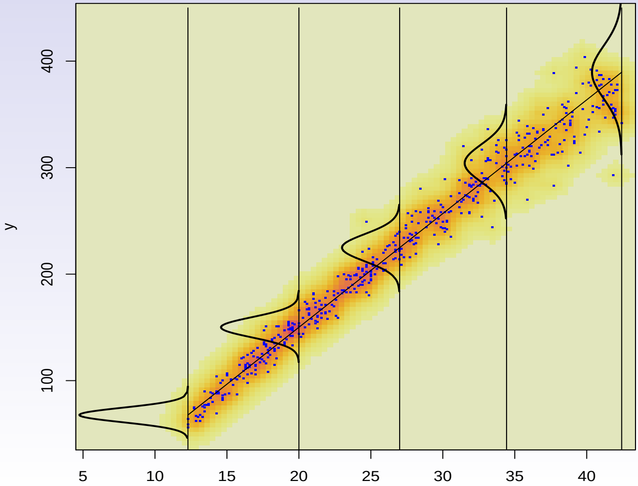
The linear model assumptions

Go back 70-80 Next page



The weighted linear model assumptions

Go back 70-80 Next page



gamlss

The linear model: comments

- estimation is achieved by Least Squares or Weighted Least Squares (WLS)
- the normal distribution is important for inference
- we only modelling the mean as linear function of the explanatory variables
- One of the top ten reasons to become statistician (according to Friedman, Friedman & Amoo, 2002, Journal of Statistics Education):

"Statisticians are mean lovers".

Go back 70-80

The generalised linear model

Generalised Linear Model, Nelder and Wedderburn(1972)

$$g(\boldsymbol{\mu}) = \mathbf{X}\beta \text{ where } \mathbf{y} \sim \mathit{ExpFamily}(\boldsymbol{\mu}, \phi)$$

where $g()$ is the link function

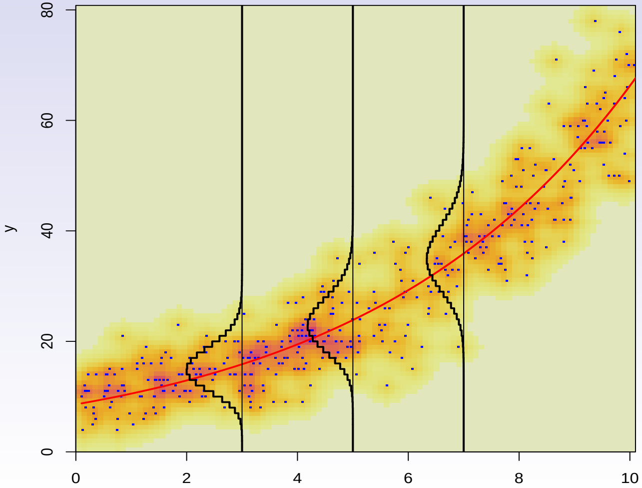
The exponential family

1. normal
2. Gamma
3. inverse Gaussian
4. Poisson
5. binomial

Go back 70-80 Next page

The generalised linear model

Go back 70-80 Next page



The generalised linear model

- estimation is achieved by Iterative Reweighted Least Squares (IRLS)
- we can model discrete response variables
- we are still "mean lovers".

Go back 70-80

gamlss

The generalised additive model

Generalised additive model Hastie and Tibshirani (1990)

$$\mathbf{y} \sim ExpFamily(\boldsymbol{\mu}, \phi)$$

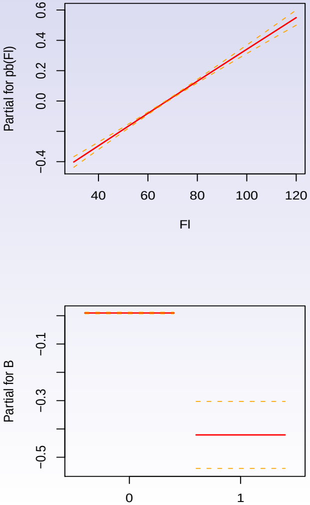
$$g(\boldsymbol{\mu}) = \mathbf{X}\beta + h_1(\mathbf{x}_1) + ... + h_k(\mathbf{x}_k)$$

where $h_j(\mathbf{x}_j)$ are smooth functions of the X's.

Go back 90 Next page

The generalised additive model

Go back 90 Next page



The generalised additive model

- the estimation is achieved by modified backfitting
- modified backfitting: is a combination of IRLS for the linear part and penalised IRLS for the smoothing part.
- the smoothing parameters λ's are fixed
- we are still "mean lovers".

Go back 90

gamlss

The Lambda, Mu, and Sigma (LMS) method

Let Y be a random variable with range $Y > 0$ defined through the transformed variable Z given by:

$$\begin{aligned} Z &= \frac{1}{\sigma\nu}\left[\left(\frac{Y}{\mu}\right)^{\nu} - 1\right], \quad \text{if} \quad \nu \neq 0 \\ &= \frac{1}{\sigma}\log\left(\frac{Y}{\mu}\right), \quad \text{if} \quad \nu = 0. \end{aligned}$$

where $z \sim N(0,1)$ (truncated) then $Y \sim BCCG(\mu, \sigma, \nu) =$ LMS method

$$\begin{aligned} \mu &= s(x, df_{\mu}) \\ \log(\sigma) &= s(x, df_{\sigma}) \\ \nu &= s(x, df_{\nu}) \end{aligned}$$

The LMS method

- modelling three parameters as smooth function of one explanatory variable
- LMS is a method of centile estimation (not a distribution)
- identity link for mu?
- smoothing parameters (df's) are fixed
- mu is the median (not mean lovers any more)
- all the parameters of the distributions are modelled as function of one explanatory variable

Go back 90

Mean And Dispersion Additive Models

Rigby and Stasinopoulos (1994, 1996a, 1996b)

$$\mathbf{y} \sim NO(\boldsymbol{\mu}, \boldsymbol{\sigma})$$

or

$$\mathbf{y} \sim EQL(\boldsymbol{\mu}, \boldsymbol{\sigma})$$

$$\begin{aligned} g_\mu(\boldsymbol{\mu}) &= \mathbf{X}_\mu \beta + h_{1,\mu}(\mathbf{x}_1) + \ldots + h_k(\mathbf{x}_{k,\mu}) \\ g_\sigma(\boldsymbol{\sigma}) &= \mathbf{X}_\sigma \beta + h_{1,\sigma}(\mathbf{x}_1) + \ldots + h_k(\mathbf{x}_{k,\sigma}) \end{aligned}$$

where EQL is the Extended Quasi Likelihood Nelder (1992) and $h_j(\mathbf{x}_j)$ are smooth functions of the X's.

Go back 90

Generalised Additive Model for Location Scale and Shape

Generalised additive model for location scale and shape Rigby and Stasinopoulos (2005)

$$\mathbf{y} \sim D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau})$$

$$\begin{aligned}
g_\mu(\boldsymbol{\mu}) &= \mathbf{X}_\mu\beta + h_{1,\mu}(\mathbf{x}_1) + \ldots + h_k(\mathbf{x}_{k,\mu}) \\
g_\sigma(\boldsymbol{\sigma}) &= \mathbf{X}_\sigma\beta + h_{1,\sigma}(\mathbf{x}_1) + \ldots + h_k(\mathbf{x}_{k,\sigma}) \\
g_\nu(\boldsymbol{\nu}) &= \mathbf{X}_\nu\beta + h_{1,\nu}(\mathbf{x}_1) + \ldots + h_k(\mathbf{x}_{k,\nu}) \\
g_\tau(\boldsymbol{\tau}) &= \mathbf{X}_\tau\beta + h_{1,\tau}(\mathbf{x}_1) + \ldots + h_k(\mathbf{x}_{k,\tau})
\end{aligned}$$

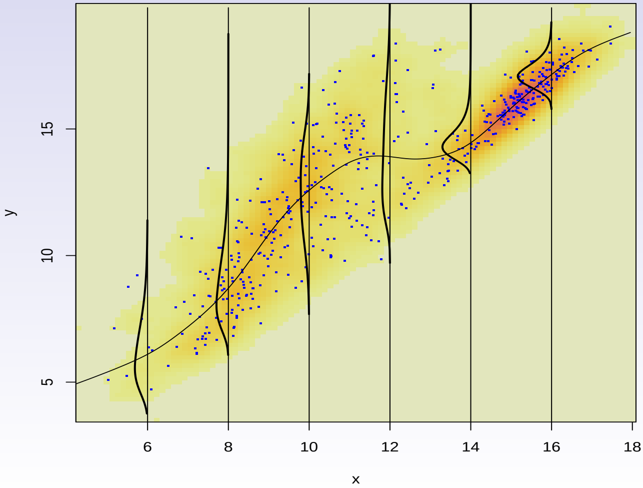
where $D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau})$ can be any distribution and where $h_j(\mathbf{x}_j)$ are smooth functions of the X's.

gamlss

Go back
next page assumptions

GAMLSS assumptions

Go back 90 Next page



gamlss

GAMLSS: comments

- Any distribution
- because of backfitting different additive terms can be implemented
- in fact any Gaussian Markov random field Rue and Held (2005)
- the algorithm maximised a penalised log-Likelihood for fixed smoothing parameters
- two algorithms go to algorithms to fit the model exist both using (penalised) IRLS and backfitting.

Go back 90

gamlss

GAMLSS: algorithms

RS a generalization of the MADAM algorithm, Rigby and Stasinopoulos (1996a)

CG a generalization of Cole and Green (1992) algorithm

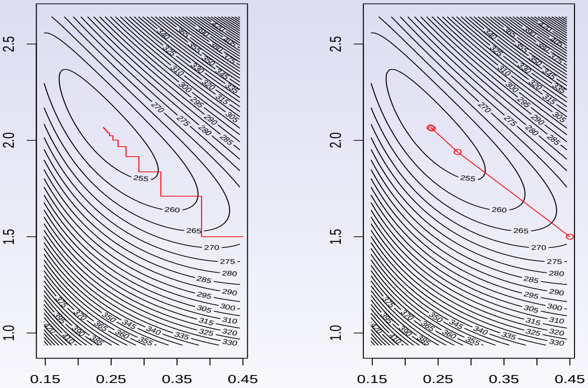
mixed a mixture of RS+CG (i.e. j iterations of RS, followed by k iterations of CG)

Go back 90 | back to GAMLSS comments | next page

gamlss

GAMLSS: algorithms

Go back 90 back to GAMLSS comments next page



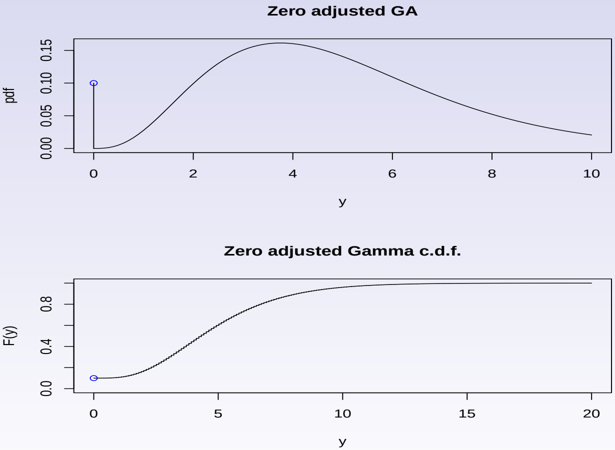
What is GAMLSS?: advantages of algorithms

1. flexible modular fitting procedure
2. easy implementation of new distributions
3. easy implementation of new additive terms
4. simple starting values for (μ, σ, ν, τ) easily found
5. stable and reliable algorithms
6. fast fitting (for fixed hyperparameters)

Go back 90 Go back Algorithm Go back GAMLSS comments

Example of mixed distribution distributions

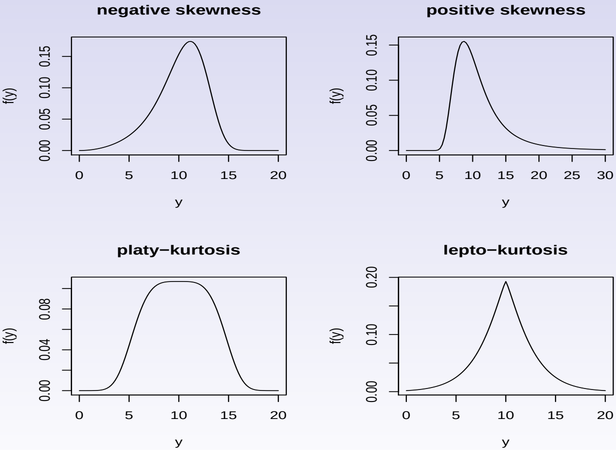
Go back Distributions



gamlss

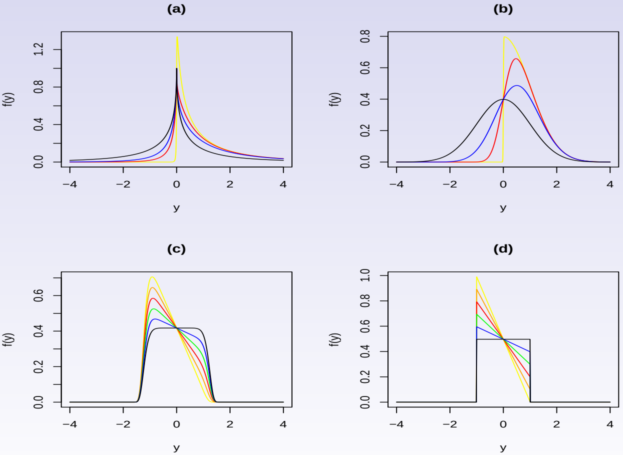
Continuous distributions: different shapes

Go back Distributions



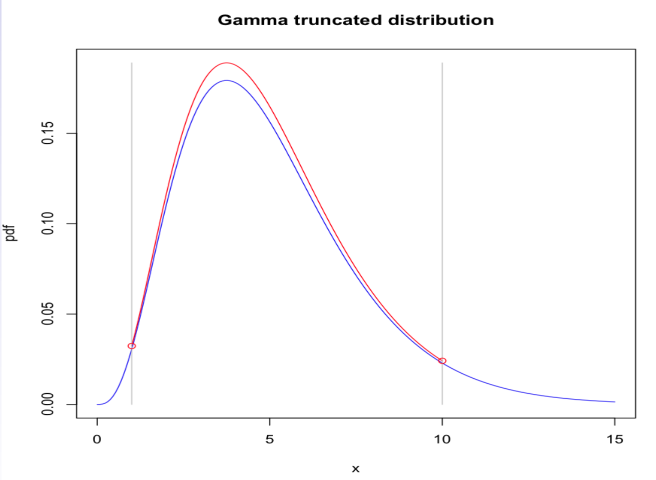
Continuous distributions: different types

Go back Distributions



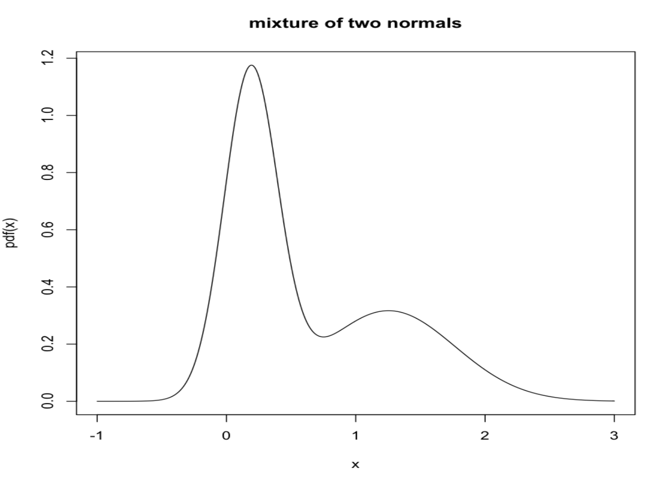
Continuous distributions: different types

Go back Distributions



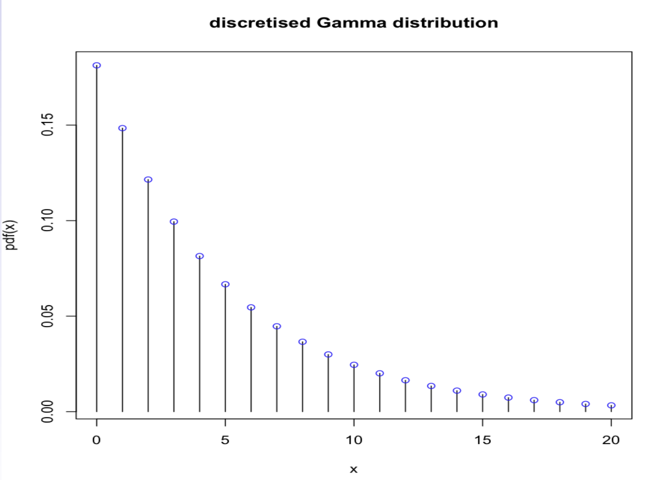
Continuous distributions: different types

Go back Distributions



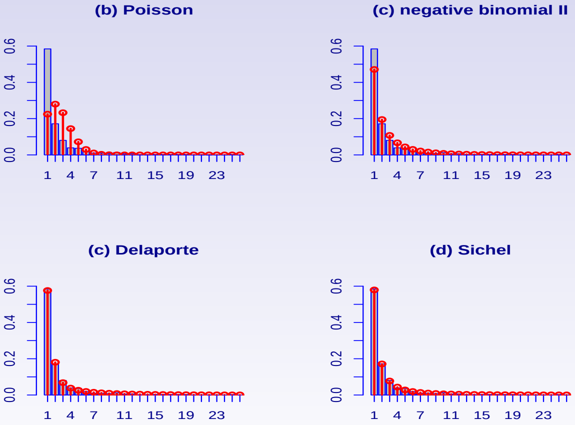
Continuous distributions: different types

Go back Distributions



gamlss

The stylometric data, Go back to distributions



gamlss

WHO

- The World Health Organisation decide in 2004-05 to adopt the BCPE distribution Rigby and Stasinopoulos (2004) and GAMLSS for the construction of Child Growth Standard curves.
- Since 2006, when the World Health Organisation launched the new Child Growth Standards, over 140 countries have adopted them.
- Three books were released:
  - WHO Multicentre Growth Reference Study Group (2006): Length, Height, weight and body mass index
  - WHO Multicentre Growth Reference Study Group (2007): Head circumference, arm circumference, triceps circumference and subscapular skinford
  - WHO Multicentre Growth Reference Study Group (2009): Growth velocity based on weight, length and head circumference. Go back to chronology

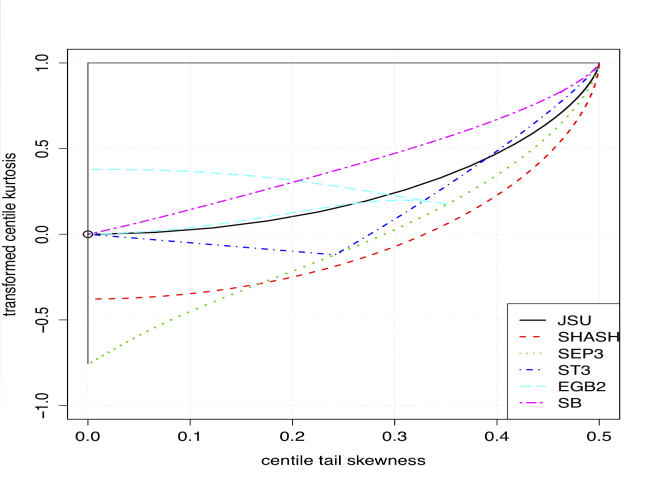
gamlss

ACEGES

- ACEGES Stands for Agent-based Computational Economics of the Global Energy System.
- People: Vlasis Voudouris, Ken'ichi Matsumoto, Konstantinos Skindilias, Bob Rigby, Paul Eilers, Michael Jefferson, John Sedgwick, Carlo Di Maio
- Interests: Oil, Gas, Electricity, Renewable energy

Go back

Centile based properties, Go back to model selection



gamlss

Question one: selection of explanatory variables

Go back to model selection

How to select explanatory variables?

	x_1	x_2	x_3	x_4	x_5	x_6
μ						
σ						
ν						
τ						

- Strategy A
- Strategy B

Go back to model selection next page

Questione one: strategy A

Go back to model selection

Strategy A:

- It starts with a forward stepwise selection using GAIC.
- Each x variables is set for selection first for μ then for σ, ν and τ
- then it does a backward elimination for ν, σ and μ.

	x_1	x_2	x_3	x_4	x_5	x_6
μ	$\times$		$\times$	$\times$		$\times$
σ			$\times$	$\times$		
ν	$\times$		$\times$			
τ				$\times$		

Go back to model selection previous page next page

Question one: strategy B

Go back to model selection

Strategy B: forward stepwise selection using GAIC in which an x variable is selected for all the parameters

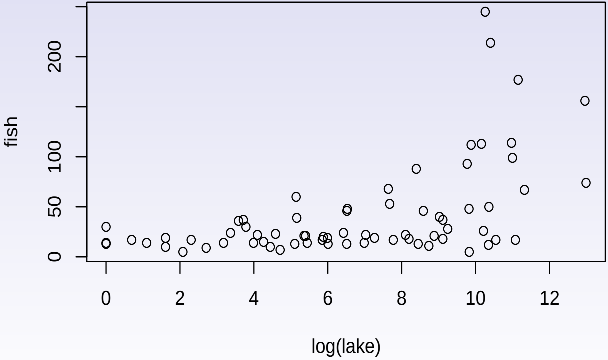
Table: selecting explanatory variables

	x_1	x_2	x_3	x_4	x_5	x_6
μ	×		×			×
σ	×		×			×
ν	×		×			×
τ	×		×			×

Go back to model selection previous page

Discrete distributions: the fish species data

Go back to present and future   next page



gamlss

Discrete distributions: the fish species data

There are several questions that need to be answered.

- How does the mean of y depend on x?
- Is y overdispersed Poisson?
- How does the variance y depend on its mean?
- What is the distribution of y given x?
- Do the scale and shape parameters of the distribution of y depend on x?

Go back to present and future next page

gamlss

Discrete distributions: Overdispersed count data approaches

Go back to present and future next page

Model	$f_Y(y)$	μ	σ	ν	DEV	df	AIC	SBC
1	PO	$x < 2 >$	-	-	1849.3	3	1855.3	1862.0
2	NBI	x	1	-	619.8	3	625.8	632.6
3	NBI	$x < 2 >$	1	-	614.3	4	622.3	631.3
4	NBI	$cs(x, 3)$	1	-	611.9	6	623.9	637.4
5	NBI	$x < 2 >$	x	-	605.0	5	615.0	626.2
6	NBI-fam	$x < 2 >$	1	1	606.0	5	616.0	627.3
7	NBI-fam	$x < 2 >$	x	1	604.9	6	616.9	630.4

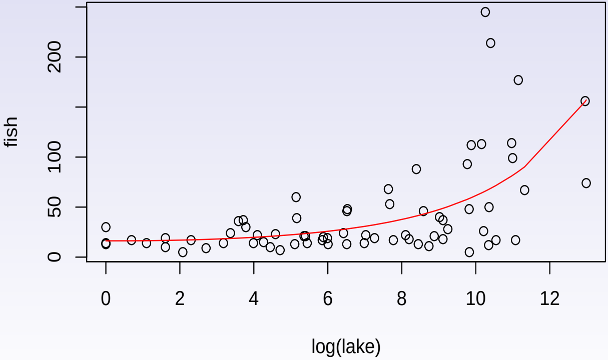
Discrete distributions: Overdispersed count data approaches

Go back to present and future next page

Model	$f_Y(y)$	μ	σ	ν	DEV	df	AIC	SBC
8	PIG	$x < 2 >$	1	-	613.3	4	621.3	630.3
9	SI	$x < 2 >$	1	x	597.7	6	609.7	623.2
10	DEL	$x < 2 >$	1	x	600.6	6	612.6	626.1
11	DEL	$x < 2 >$	-	x	600.6	5	610.6	621.9
12	PO-Normal	$x < 2 >$	1	-	615.2	4	623.2	632.2
13	NBI-Normal	$x < 2 >$	x	1	603.7	6	615.7	629.2
14	PO-NPFM(5)	$x < 2 >$	-	—	601.9	13	627.9	657.2
15	NB-NPFM(2)	$x < 2 >$	1	—	611.9	6	623.9	637.4
16	doublePO	$x < 2 >$	x	-	616.4	5	626.4	637.6
17	IGdisc	$x < 2 >$	1	-	603.3	4	611.3	620.3

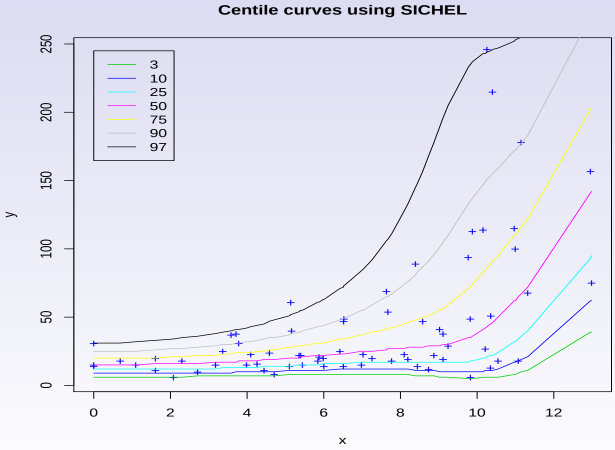
Discrete distributions: fitted mean of the Sichel distribution

Go back to present and future next page



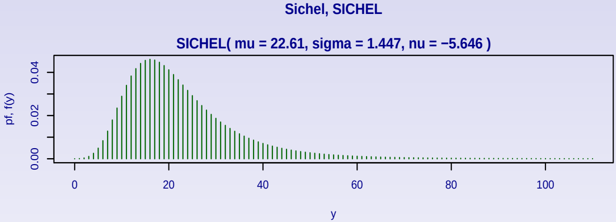
Discrete distributions: fitted centiles

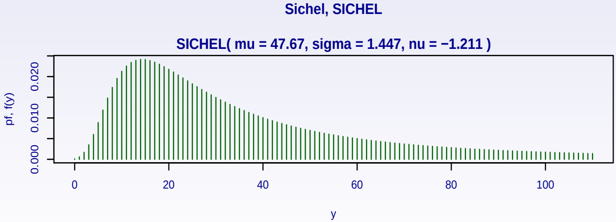
Go back to present and future   next page



Discrete distributions: fitted Sichel distributions for observations (a) 40 and (b) 67

Go back to present and future





GAMLSS impact

- GAMLSS is used in different applied scientific fields
  - personal research Go to personal research
  - other people's research Go to others'
- more than 250 Google scholar citations
- more that 5000 hits in the GAMLSS site per year Go to hits

Go back to chronology

gamlss

GAMLSS our personal research

- Centile Estimation: Anthony L. Guerrerio, Duarte Freitas, Daniel Cohen, Gavin Sandercock, David McCarthy
- Medical applications: Graciela Muniz Terrera, Ardo Van Den Hout,Stefan van Buuren, Daniel J. Hayes, D., Feiko O. ter Kuile, Dianne J. Terlouw
- Insurance data: Gillian Heller
- River flow data: Floris van Ogtrop, Gillian Heller
- Film data Vlasios Voudouris, Bob Gilchrist, Bob Rigby, John Sedgwick
- Financial data Nikolaos Georgikopoulos, Timemaxos Efthimiades
- Distributions Popi Akantziliotou, Fiona McElduff
- Oil and Gas data Vlasios Voudouris, Kenichi Matsumoto, Majid Djennad,

GAMLSS: Survival analysis

Go back impact

ELSEVIER

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

A hands-on approach for fitting long-term survival models under the GAMLSS framework

Mário de Castro[a,*], *Vicente G. Cancho*[a], *Josemar Rodrigues*[b]

[a] *Universidade de São Paulo, Instituto de Ciências Matemáticas e de Computação, Caixa Postal 668, 13560-970, São Carlos-SP, Brazil*
[b] *Universidade Federal de São Carlos, Departamento de Estatística, Via Washington Luís, km 235, Caixa Postal 676, 13565-905, São Carlos-SP, Brazil*

ARTICLE INFO

Article history:
Received 19 October 2008
Received in revised form
5 August 2009
Accepted 11 August 2009

Keywords:
Survival analysis
Cure rate models
Long-term survival models
GAMLSS
Negative binomial distribution

ABSTRACT

In many data sets from clinical studies there are patients insusceptible to the occurrence of the event of interest. Survival models which ignore this fact are generally inadequate. The main goal of this paper is to describe an application of the generalized additive models for location, scale, and shape (GAMLSS) framework to the fitting of long-term survival models. In this work the number of competing causes of the event of interest follows the negative binomial distribution. In this way, some well known models found in the literature are characterized as particular cases of our proposal. The model is conveniently parameterized in terms of the cured fraction, which is then linked to covariates. We explore the use of the `gamlss` package in R as a powerful tool for inference in long-term survival models. The procedure is illustrated with a numerical example.

gamlss

GAMLSS: Boosting

Go back impact

GAMLSS for high-dimensional data – a flexible approach based on boosting

Andreas Mayr*[1], Nora Fenske[2], Benjamin Hofner[1], Thomas Kneib[3], Matthias Schmid[1]

[1] Institut für Medizininformatik, Biometrie und Epidemiologie, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany
[2] Institut für Statistik, Ludwig-Maximilians-Universität München, Germany
[3] Institut für Mathematik, Carl von Ossietzky Universität Oldenburg, Germany

GAMLSS: Deer

Go back impact

Body-mass or sex-biased tick parasitism in roe deer (*Capreolus capreolus*)? A GAMLSS approach

C. KIFFNER, C. LÖDIGE, M. ALINGS, T. VOR and F. RÜHE
Department of Forest Zoology and Forest Conservation incl. Wildlife Biology and Game Management, Büsgen-Institute, Georg-August-University Göttingen, Göttingen, Germany

Abstract. Macroparasites feeding on wildlife hosts follow skewed distributions for which basic statistical approaches are of limited use. To predict *Ixodes* spp. tick burden on roe deer, we applied Generalized Additive Models for Location, Scale and Shape (GAMLSS) which allow incorporating a variable dispersion. We analysed tick burden of 78 roe deer, sampled in a forest region of Germany over a period of 20 months. Assuming a negative binomial error distribution and controlling for ambient temperature, we analysed whether host sex and body mass affected individual tick burdens. Models for larval and nymphal tick burden included host sex, with male hosts being more heavily infested than female ones. However, the influence of host sex on immature tick burden was associated with wide standard errors (nymphs) or the factor was marginally significant (larvae). Adult tick burden was positively correlated with host body mass. Thus, controlled for host body mass and ambient temperature, there is weak support for sex-biased parasitism in this system. Compared with models which assume linear relationships, GAMLSS provided a better fit. Adding a variable dispersion term improved only one of the four models. Yet, the potential of modelling dispersion as a function of variables appears promising for larger datasets.

Key words. Cervidae, *Ixodes ricinus*, aggregation, ecto-parasite, parasitism, seasonality, sex-bias, tick-borne disease.

GAMLSS: MSc thesis

Go back impact

ENGENHARIA DE AVALIAÇÕES COM BASE EM

MODELOS GAMLSS

LUTEMBERG DE ARAÚJO FLORENCIO

Orientador: Prof. Dr. Francisco Cribari Neto

Co-orientador: Prof. Dr. Raydonal Ospina Martínez

Área de Concentração: Estatística Aplicada

GAMLSS: Go back impact

Uma aplicação do modelo GAMLSS paramétrico a dados de seguro de saúde

Gustavo H. A. Pereira   Jalmar M. F. Carrasco
Departamento de Estatística, Instituto de Matemática e Estatística, Universidade de São Paulo

1 Introdução

GAMLSS: Eucalypt flowering

Go back impact

Climate effects and temperature thresholds for Eucalypt flowering: a GAMLSS ZIP approach

I.L. Hudson[a], S.W. Kim[b], and M.R. Keatley[c]

[a] *School of Mathematical and Physical Sciences, University of Newcastle, NSW, Australia*
[b] *Flinders Centre for Epidemiology and Biostatistics, Flinders University, Adelaide, South Australia*
[c] *Dept of Forest and Ecosystem Science, The University of Melbourne, Creswick, Victoria, Australia*
Email: Irene.Hudson@newcastle.edu.au

GAMLSS: Electricity prices

Go back impact

Distributional modeling and short-term forecasting of electricity prices by Generalized Additive Models for Location, Scale and Shape

Francesco Serinaldi *

School of Civil Engineering and Geosciences, Newcastle University, Newcastle Upon Tyne, NE1 7RU, UK
Willis Research Network, London, UK

A R T I C L E   I N F O

Article history:
Received 21 March 2010
Received in revised form 10 February 2011
Accepted 1 May 2011
Available online 8 May 2011

JEL classification:
C01
C18
C32
C52
C53
Q40

Keywords:
Electricity market price
Point forecasting
Interval forecasting
GAMLSS models
Liberalized energy markets

A B S T R A C T

In the context of the liberalized and deregulated electricity markets, price forecasting has become increasingly important for energy company's plans and market strategies. Within the class of the time series models that are used to perform price forecasting, the subclasses of methods based on stochastic time series and causal models commonly provide point forecasts, whereas the corresponding uncertainty is quantified by approximate or simulation-based confidence intervals. Aiming to improve the uncertainty assessment, this study introduces the Generalized Additive Models for Location, Scale and Shape (GAMLSS) to model the dynamically varying distribution of prices. The GAMLSS allow fitting a variety of distributions whose parameters change according to covariates via a number of linear and nonlinear relationships. In this way, price periodicities, trends and abrupt changes characterizing both the position parameter (linked to the expected value of prices), and the scale and shape parameters (related to price volatility, skewness, and kurtosis) can be explicitly incorporated in the model setup. Relying on the past behavior of the prices and exogenous variables, the GAMLSS enable the short-term (one-day ahead) forecast of the entire distribution of prices. The approach was tested on two datasets from the widely studied California Power Exchange (CalPX) market, and the less mature Italian Power Exchange (IPEX). CalPX data allow comparing the GAMLSS forecasting performance with published results obtained by different models. The study points out that the GAMLSS framework can be a flexible alternative to several linear and nonlinear stochastic models.

GAMLSS: Real Estate

Go back impact

A NEW METHODOLOGY FOR REAL ESTATE APPRAISAL USING GAMLSS MODELS

LUTEMBERG FLORENCIO, FRANCISCO CRIBARI-NETO, AND RAYDONAL OSPINA

ABSTRACT. The valuation of real estates (e.g., house, land, among others) is of extreme importance for decision making. Their singular characteristics make valuation through hedonic pricing methods difficult since the theory does not specify the correct regression functional form nor which explanatory variables should be included in the hedonic equation. In this article we perform real estate appraisal using a class of regression models proposed by Rigby & Stasinopoulos (2005): generalized additive models for location, scale and shape (GAMLSS). Our empirical analysis shows that these models semm to be more appropriate for estimation of the hedonic prices function than the regression models currently used to that end.

gamlss

GAMLSS: Stock Liquidity

Go back impact

《Journal of Shanxi Finance and Economics University》 2011-04

Add to Favorite Get Latest Update

Non-parametric and Non-linear Analysis of Stock Liquidity with High Frequency Data in China on GAMLSS Model

LIU Hao,CHEN Lang-nan (Lingnan College,Sun Yat-Sen University,Guangzhou 510275,China)

Empirical studies of financial time-series data show that there is strong evidence of serial correlation,time-varying heteroscedasticity and non-normal distribution.Using GAMLSS model from biostatistics,the authors adopt the BCT distribution and different non-linear non parametric models to fit high-frequency stock liquidity.Results show that BCT distribution with non-linear cubic-splines(CS) model is superior under the GAIC criteria.Adding dependent variables will increase GAMLSS's goodness of fit.

【Key Words】： **high-frequency liquidity BCT distribution GAMLSS non-parametric cubic-splines model GAIC criteria**

【Fund】：国家自然科学基金项目(70673116);; 北京大学汇丰金融研究院2009年课题;; 中山大学"985工程"产业与区域发展研究创新基地资助课题;; 国家社科基金重点课题(08ATL007);; 广东省自然科学基金课题(9151027501000032);; 广东省社科基金课题;; 广东省普通高校人文社会科学重点研究基地资助课题

【CateGory Index】： F830.91;F224

【DOI】： CNKI:SUN:SXCJ.0.2011-04-005

Download(CAJ format) Download(PDF format)

CAJViewer7.0 supports all the CNKI file formats; AdobeReader only supports the PDF format.

—【Citations】

gamlss

GAMLSS: in chinese

Go back impact

DOI:10.3969/j.issn.1673-5501.2011.05.006

应用 GAMLSS 技术构建基于性别、年龄和身高的新疆 7~17 岁儿童青少年血压参考标准

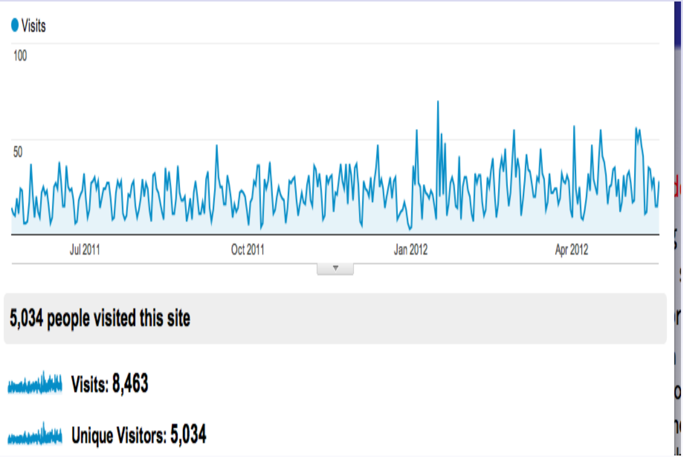
严 恺[1,2] 王 倩[2] 姚 华[2] 黄永迪[2] 王琛琛[2] 张卫国[2] 娜尔克孜·阿布扎力汗[2] 李南方[3] 严卫丽[1]

摘要 目的 结合性别、年龄和不同身高百分位数,应用 GAMLSS 技术探索性构建新疆 7~17 岁儿童青少年血压百分分

gamlss

GAMLSS: Number of hits on the website

Go back to impact



GAMLSS: Area of origin

Go back to impact

	Sub Continent Region	Visits	Page
1.	Northern America	1,953	
2.	Northern Europe	1,716	
3.	Western Europe	1,462	
4.	South America	1,173	
5.	Southern Europe	651	
6.	Eastern Asia	419	
7.	Australia and New Zealand	325	
8.	Eastern Europe	189	
9.	South-Eastern Asia	139	
10.	Western Asia	136	

gamlss

GAMLSS: The read paper

Go back

Appl. Statist. (2005)
54, *Part* 3, *pp.* 507–554

Generalized additive models for location, scale and shape

R. A. Rigby and D. M. Stasinopoulos
London Metropolitan University, UK

[*Read before* The Royal Statistical Society *on Tuesday, November 23rd, 2004, the President*, Professor A. P. Grieve, *in the Chair*]

Summary. A general class of statistical models for a univariate response variable is presented which we call the generalized additive model for location, scale and shape (GAMLSS). The model assumes independent observations of the response variable *y* given the parameters, the explanatory variables and the values of the random effects. The distribution for the response variable in the GAMLSS can be selected from a very general family of distributions including highly skew or kurtotic continuous and discrete distributions. The systematic part of the model is expanded to allow modelling not only of the mean (or location) but also of the other parameters of the distribution of *y*, as parametric and/or additive nonparametric (smooth) functions of explanatory variables and/or random-effects terms. Maximum (penalized) likelihood estimation is used to fit the (non)parametric models. A Newton–Raphson or Fisher scoring algorithm is used to maximize the (penalized) likelihood. The additive terms in the model are fitted by using a backfitting algorithm. Censored data are easily incorporated into the framework. Five data sets from different fields of application are analysed to emphasize the generality of the GAMLSS class of models.

Keywords: Beta–binomial distribution; Box–Cox transformation; Centile estimation; Cubic smoothing splines; Generalized linear mixed model; *LMS* method; Negative binomial distribution; Non-normality; Nonparametric models; Overdispersion; Penalized likelihood; Random effects; Skewness and kurtosis

gamlss

Mikis in RSS

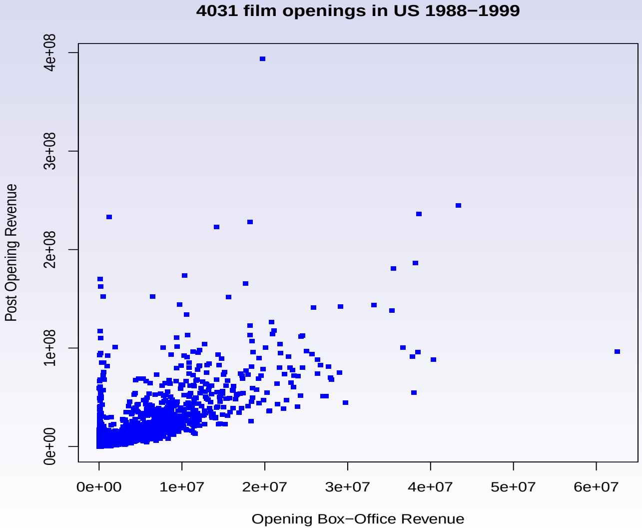
Go back

Bob in RSS

Go back

Film Data

Go back



gamlss

Film Data: the quiz

A small Quiz

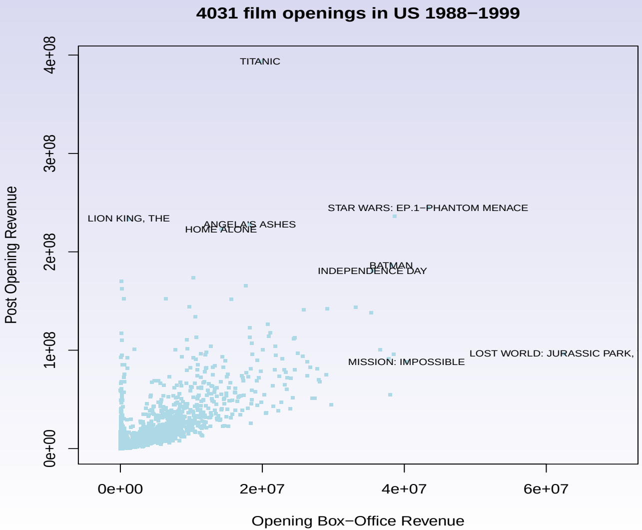
- Question 1 which film in the 90's has made the most money?
- Question 2 which film in the 90's did badly in the first week but subsequently did very well?
- Question 3 which film in the 90's did well in the first week but flopped after that?

Go back

gamlss

Film Data: Answers to the quiz

Go back



gamlss

The GEST Process: definition

Go back to conclusions next page

GEST process

$$\begin{aligned}
Y_t|\mu_t, \sigma_t, \nu_t, \tau_t &\sim D(\mu_t, \sigma_t, \nu_t, \tau_t) \\
g_1(\mu_t) = \gamma_{1,t} &= \sum_{j=1}^{J_1} \phi_{1,j}\gamma_{1,t-j} + b_{1,t} \\
g_2(\sigma_t) = \gamma_{2,t} &= \sum_{j=1}^{J_2} \phi_{1,j}\gamma_{2,t-j} + b_{2,t} \\
g_3(\nu_t) = \gamma_{3,t} &= \sum_{j=1}^{J_3} \phi_{3,j}\gamma_{3,t-j} + b_{3,t} \\
g_4(\tau_t) = \gamma_{4,t} &= \sum_{j=1}^{J_4} \phi_{4,j}\gamma_{4,t-j} + b_{4,t}
\end{aligned}$$

where $b_{k,t} \sim N(0, \sigma_{b,k})$

The GEST Process: SST Process

Go back to conclusions next page

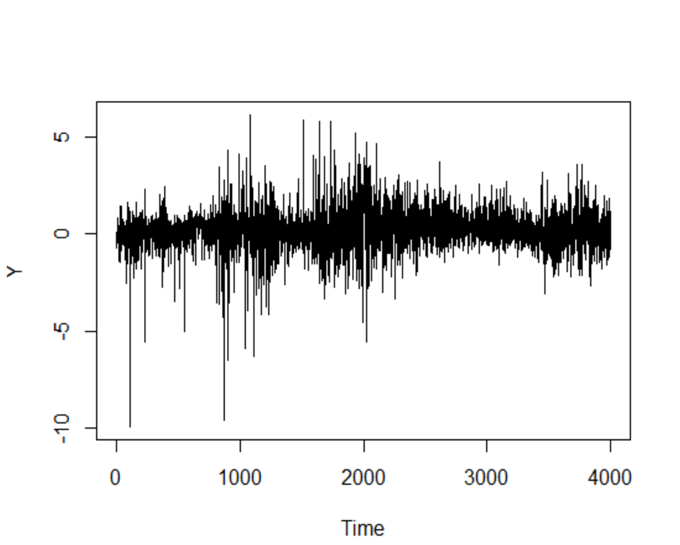
Skew Student *t* distribution

$$\begin{aligned}
Y_t|\mu_t, \sigma_t, \nu_t, \tau_t &\sim SST(\mu_t, \sigma_t, \nu_t, \tau_t) \\
\mu_t &= \mu_{t-1} + b_{1,t} \\
\log(\sigma_t) &= \log(\sigma_{t-1}) + b_{2,t} \\
\log(\nu_t) &= \log(\nu_{t-1}) + b_{3,t} \\
\log(\tau_t - 2) &= \log(\tau_{t-1} - 2) + b_{4,t},
\end{aligned}$$

gamlss

GEST Process: example

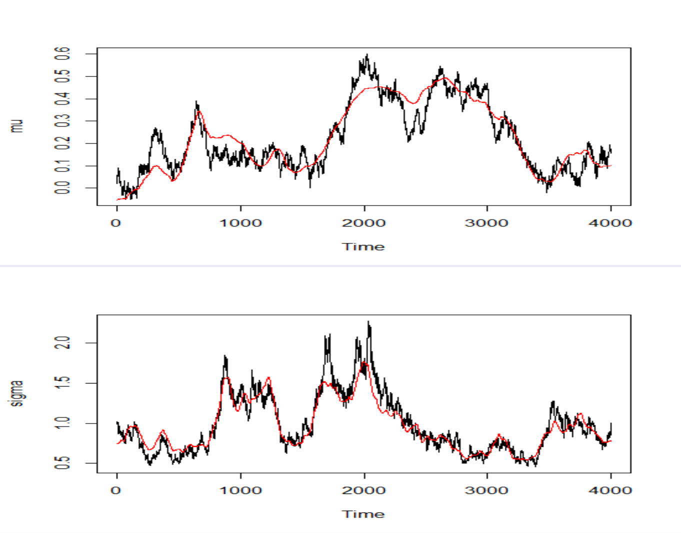
Go back to conclusions next page



gamlss

GEST Process: example

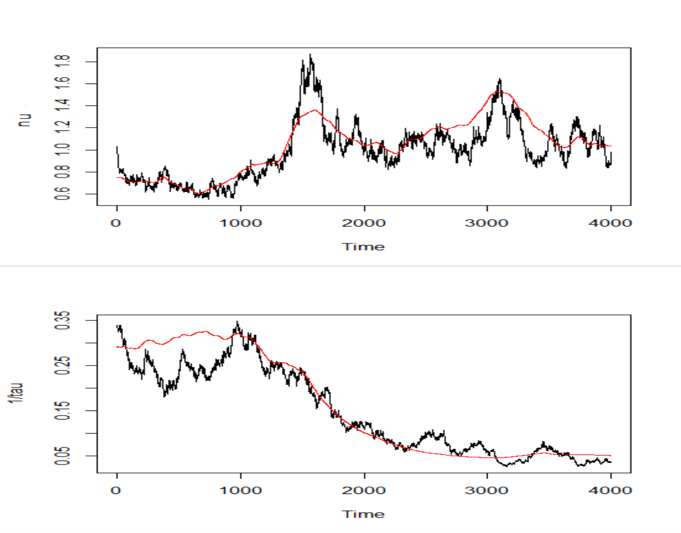
Go back to conclusions next page



gamlss

GEST Process: example

Go back to conclusions



gamlss

Boosting

Go back to conclusions

GAMLSS for high-dimensional data – a flexible approach based on boosting

Andreas Mayr*[1], Nora Fenske[2], Benjamin Hofner[1], Thomas Kneib[3], Matthias Schmid[1]

[1] Institut für Medizininformatik, Biometrie und Epidemiologie, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany
[2] Institut für Statistik, Ludwig-Maximilians-Universität München, Germany
[3] Institut für Mathematik, Carl von Ossietzky Universität Oldenburg, Germany

Modelling the tail

Go back to model selection